UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION III
1650 Arch Street
Philadelphia, Pennsylvania 19103-2029

Mr. John Wirts
Division of Water and Waste Management
West Virginia Department of Environmental Protection
601 57th Street SE
Charleston, WV 25304

      Re: W.Va. Code §22-11-7b(f) Assessment Methodology for the Biological Component of the
      Narrative Criteria in Wadeable Streams

Dear Mr. Wirts:

Thank you for the opportunity to comment on the recently proposed procedural rule: *Assessment Methodology for the Biological Component of the Narrative Criteria in Wadeable Streams* filed on February 9, 2021. The U.S. Environmental Protection Agency (EPA) has reviewed the proposed procedural rule and is providing the below technical comments to the West Virginia Department of Environmental Protection (WVDEP) for consideration. EPA commends WV for continuing to develop and maintain a robust water quality monitoring program, and for its commitment to use the data that is collected for listing and assessment purposes. However, based on EPA's current analysis, it appears that WVDEP's proposed bioassessment methodology may result in a substantial number of West Virginia waters being inaccurately assessed as unimpaired, thus leaving those waters without the protections they would otherwise be afforded under the Clean Water Act (i.e., Total Maximum Daily Loads and/or more stringent effluent limits for permitted discharges). Below, I have included general and more specific comments to make certain EPA fulfills its obligation to ensure state 303(d) lists of impaired waters meet all applicable federal regulations at 40 Code of Federal Regulations (CFR) § 130.7(b).

General Comments:
- It is a best practice and an EPA recommendation to routinely recalibrate an index of biological integrity (IBI) like the West Virginia Stream Condition Index (WVSCI) with updated data to ensure it will provide accurate results. The scoring criteria used for WVSCI and its component metrics have not been updated or recalibrated since 2000. WVDEP has collected bioassessment data at hundreds of additional reference sites since 2000 which should be incorporated into West Virginia's IBI. Further, WVSCI uses family-level macroinvertebrate data when WVDEP's genus-level macroinvertebrate data would provide higher resolution information regarding the aquatic community.
- EPA commends WVDEP in the use of genus-level observed/expected (O/E) models to assess stressors. In order to allow EPA to conduct a complete review, please provide a copy of the *Tetra Tech Memo: Methods & Results of Site-Specific Biological Modeling (O/E) with Stressor Module Task* (Feb. 26, 2019) that is listed as a reference in the *Aquatic Life Use Assessment and Biological Stressor Identification Procedures*.

- EPA suggests WVDEP consider an independent third party bioassessment program review to identify areas where WVDEP can further enhance its listing and assessment protocols. The Midwest Biodiversity Institute has experience evaluating approximately 20 state bioassessment programs and is available to support WVDEP. If WVDEP is interested in a third-party assessment, EPA can assist in this effort.
- EPA recommends WVDEP consider whether the proposed procedural rule would comprise a revised water quality standard that must be submitted to EPA for review under section 303(c) of the Clean Water Act.

Taxonomy:
- WVDEP has been collecting genus level macroinvertebrate samples for 20 years. These data provide higher resolution information than family level data. Based on EPA's analysis of WVDEP's bioassessment database, it appears that a substantial number of West Virginia waters will be left off of West Virginia's 303(d) list of impaired waters (for which TMDLs must be developed) if WVDEP does not evaluate its complete genus level macroinvertebrate dataset. In order to meet regulatory requirements to evaluate all existing and readily available data at CFR 130.7(b)(5), EPA recommends that WVDEP use a method that evaluates its genus-level data. One such method is the *Genus Level Index of Most Probable Stream Status* (GLIMPSS).
- The use of genus level biological data is supported by several studies, including a recent analysis performed by Dr. Ryan King from Baylor University, whose review of WVDEP data from the Mountain Summer stratum using Threshold Indicator Taxa Analysis (TITAN) demonstrated that the proposed WVSCI thresholds allow significant loss (reduced frequency and abundance) of sensitive genera. For example, 80% of sensitive genera experienced significant losses (reduced frequency and abundance) at a WVSCI threshold of 72. Below a WVSCI of 61, virtually no sensitive genera remained. The same analysis also revealed that at the family level, 65% of sensitive families experienced significant loss below a WVSCI score of 80. Dr. King presented his findings at the 2019 Society for Freshwater Science Annual Meeting and at WVDEP's April 20, 2020 public hearing on a previously proposed bioassessment rule. An electronic copy of Dr. King's presentation is included with EPA's comments for reference (see Enclosure 1).

Scoring Thresholds:
- The technical materials available to review for this proposed rule do not include any statistical or scientific support for the proposed WVSCI impairment threshold of 50. Please provide the supporting information and/or analysis that WVDEP relied on to establish this threshold. Biological impairment thresholds should be derived through statistical means that track stressor responses or relate to a distribution of reference sites.
- The proposal provides no scientific basis for requiring additional information (either evaluation or additional samples) for sites that score between 50 and 72. In past Integrated Report cycles, WVDEP has identified ranges of scores where WVDEP has required additional data to account for uncertainty. EPA previously has advised that WVDEP's methods for selecting these types of scoring thresholds are statistically unsupported because statical error is already accounted for in the underlying reference data on which the attainment threshold is based. To the extent WVDEP desires to identify a range of scores to account for statistical error, EPA continues to recommend that WVDEP adopt an appropriate statistical method for deriving it (such as the interval/equivalence statistical test (Kilgour et. al. 1998)). In 2015, EPA presented the interval/equivalence method to WVDEP and offered assistance in statistical calculations and

interpretation. EPA staff remain available to provide technical support to WVDEP to ensure WVDEP's final rule is statistically valid.

- No justification on threshold derivations has been provided for this proposed rule. The selected thresholds do not appear to be based upon standard derivation methods and appear likely to result in numerous instances of "false negatives" (failure to detect impairment). EPA used WVDEP data (n=7650) to calculate what proportion of samples collected in locations considered to be stressed (n=1724) fell within each category of WVDEP's proposed rule. EPA found that with the proposed thresholds, only 31% of known stressed sites (n=551 out of 1724) would automatically be deemed impaired (WVSCI <50); 27% (n=469 out of 1724) would require additional stressor identification or a second biological sample (WVSCI 50-61); and 24% (n=441 out of 1724) would be placed into a scoring threshold range requiring a second biological sample (WVSCI 62-72) (see WVSCI distributions between reference and stressed sites in the figure below). Based on this information, EPA recommends that samples with a WVSCI score less than the attainment threshold (72, as proposed) should automatically be considered impaired (for further discussion on this point, see comments on "Second Sample Requirement", below).
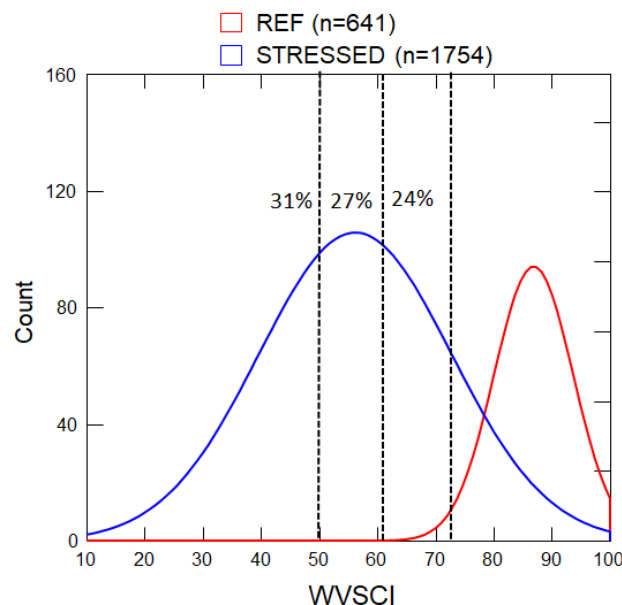

Figure 1: Distribution of WVSCI sample scores collected in stressed sites (blue line) and reference sites (red line). The three dashed lines represent different proposed categories by WVDEP. A single sample below 50 is required for an impairment listing. Between 50 and 61, additional data is required for an impairment listing, and between 61 and 72, additional biological samples are required for an impairment listing. The percentages represent the proportion of stressed site samples in each of the WVDEP categories.

- WVDEP reports the 5th percentile of the reference distribution as 72. While this may have been accurate when the threshold calculation was last performed in 2011, it no longer incorporates all available reference site data, nor does it include any updated metric standardizations that have been performed by WVDEP. In 2018, with nearly 200 additional reference sites, WVDEP calculated the 5th percentile at 74.4. While EPA recommends WVDEP utilize a tool that incorporates WVDEP's genus level macroinvertebrate data to make attainment decisions, if

WVSCI continues to be WVDEP's preferred tool, the most current WVSCI impairment threshold of 74.4 should be used. Further, the proposed method should allow for future refinement as more reference sites are identified.

Partitioning:
- EPA recommends seasonal and regional partitioning of the proposed methodology to consider equity across the state in assessment decisions. The proposed methodology using WVSCI was developed based on statewide rivers and streams encompassing broad ecoregional and seasonal scales. The broad statewide spatial scope and unadjusted seasonal application of WVSCI makes the assessment methodology less effective. EPA and WVDEP staff have noted, for example, that many Plateau streams sampled in the summer are not readily comparable to Mountain streams sampled in the spring. Seasonal and regional partitioning, for example, allows a spring Plateau sample to be compared to reference conditions in the Plateau during the spring.

Second Sample Requirement:
- The proposed method requires two samples to make an attainment determination in all circumstances except where the score is below the impairment threshold of 50, above the attainment threshold of 72, or between 50-61 with clear evidence of impairment caused by water quality or physical habitat. As noted above, if the reasoning behind taking a second sample is to account for potential sample variability, this is already accounted for by variability in the reference sites because the single sample score is compared to the reference distribution of index scores, thus making the second sample unnecessary.
- Biological monitoring is designed and calibrated to capture accurate information to make assessment decisions with one macroinvertebrate sample. Given that the calculated mean coefficient of variation of WVSCI duplicate samples is approximately 7% (less than duplicate water chemistry samples), EPA maintains that a single benthic sample provides very high confidence for assessment purposes. Use of the 5th percentile as the threshold for attainment also provides a conservative approach that reduces the likelihood of a "false positive" score (i.e., one that incorrectly identifies a sample as impaired) below the attainment threshold.
- WVDEP's proposed methodology appears to require more time and resources to assess stream condition. For example, in a two-year reporting cycle (e.g., 2016-2017 data set, n=722), EPA notes that 180 locations (25%) would require a second sample under the proposed scoring threshold range of 61-72. Based on a conservative estimate of $500/site, this requirement could potentially cost WVDEP an additional $100,000 per reporting cycle. Additional costs would also accrue in that nearly 120 samples (16%) that score in the 50-61 zone and would require resources to perform site-specific stressor identification analysis for each assessment site.
- No rational has been provided for the proposal's assumption that the second sample will be more representative than the initial sample. To the extent the goal is to account for perceived uncertainty in the sampling, WVDEP could use available data and account for seasonal and ecoregion variability by relying on the sample taken under the most limiting conditions rather than the most recent score.

Supporting Documentation:
- As noted above, the public notice did not contain technical documents to support the proposed impairment thresholds or the *Aquatic Life Use Assessment and Biological Stressor Identification Procedures* for stressor thresholds. Please provide technical support documents or further information on how thresholds were derived with expanded justifications based on the results of the statistical analyses used (listed in references as "WVDEP WAB Data Analysis. 2020").

Technical support documents would assist EPA in determining if the proposed methodology is scientifically sound. EPA requested technical support documents in our comments on previous versions of the procedural rule and none have been provided.

EPA looks forward to working with WVDEP to develop assessment procedures that will facilitate the Department's efforts to meet state and federal regulations related to water quality. If you have any questions or need any clarification on the recommendations provided above, please feel free to contact me, or have your staff contact William Richardson at 215-814-5675 or richardson.william@epa.gov.

Sincerely,

Gregory Voigt, Chief
Standards and TMDLs Section

Enclosure

Enclosure 1

Translating multimetric index scores into taxa-specific biodiversity losses

Ryan S. King[1], Emily H. Bernhardt[2], and Matthew E. Baker[3]

[1]*Department of Biology, Center for Reservoir and Aquatic Systems Research, Baylor University, Waco, TX*
www.baylor.edu/aquaticlab

[3]*Department of Biology, Duke University, Durham, NC*

[2]*Department of Geography and Environmental Systems, UMBC, Baltimore, MD*

20 May 2019
Society for Freshwater Science Annual Meeting

Multimetric indices (MMIs)

- Compare "reference" sites vs. "stressed" sites

- Identify "metrics" (aggregates) that distinguish reference and stressed sites
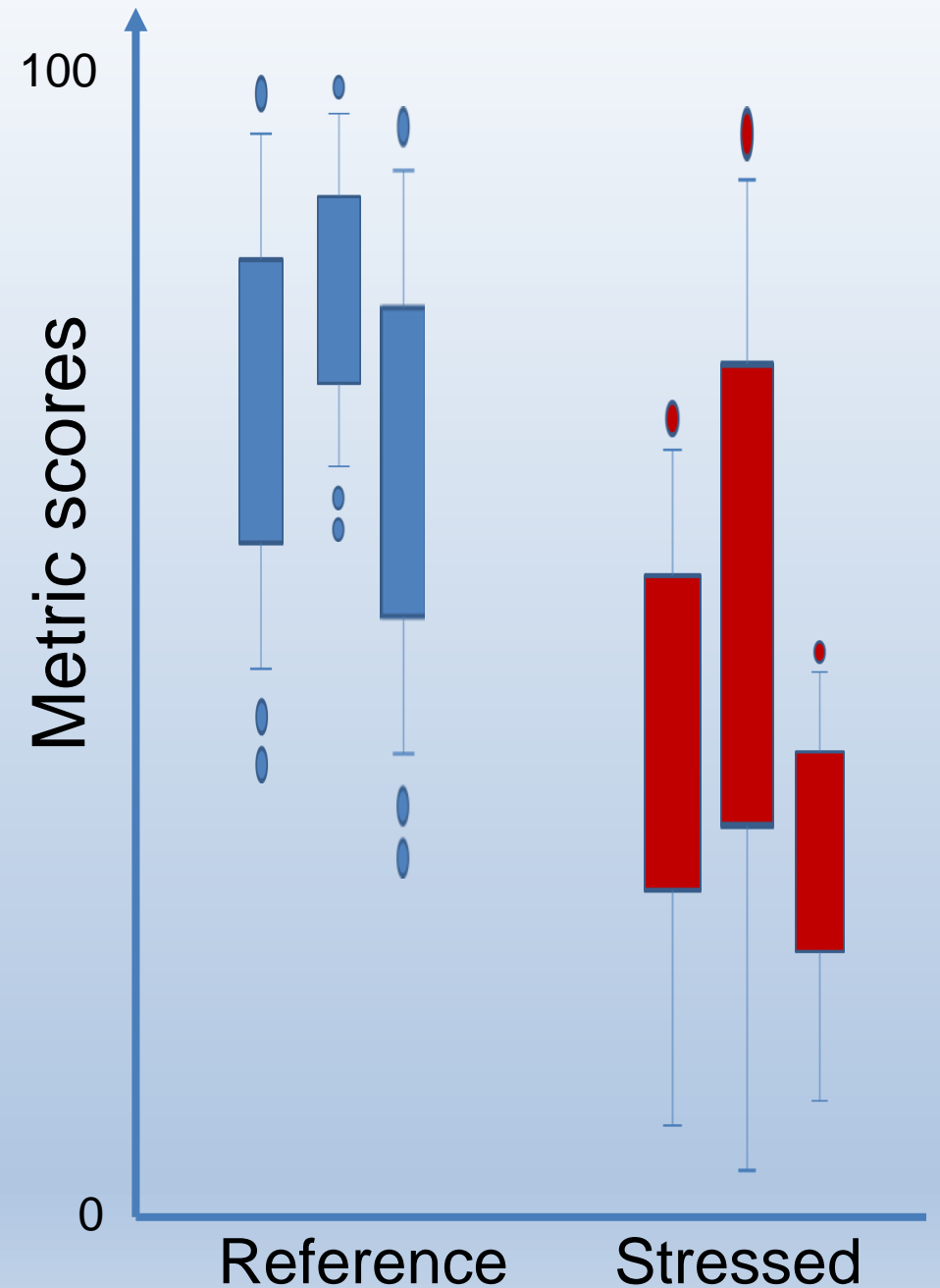
Multimetric indices (MMIs)

- Compare "reference" sites vs. "stressed" sites

- Identify "metrics" (aggregates) that distinguish reference and stressed sites

- Aggregate metric scores into a single value, i.e., MMI

- Use distribution (e.g., 5th percentile) of MMI scores in reference sites to define impairment threshold score
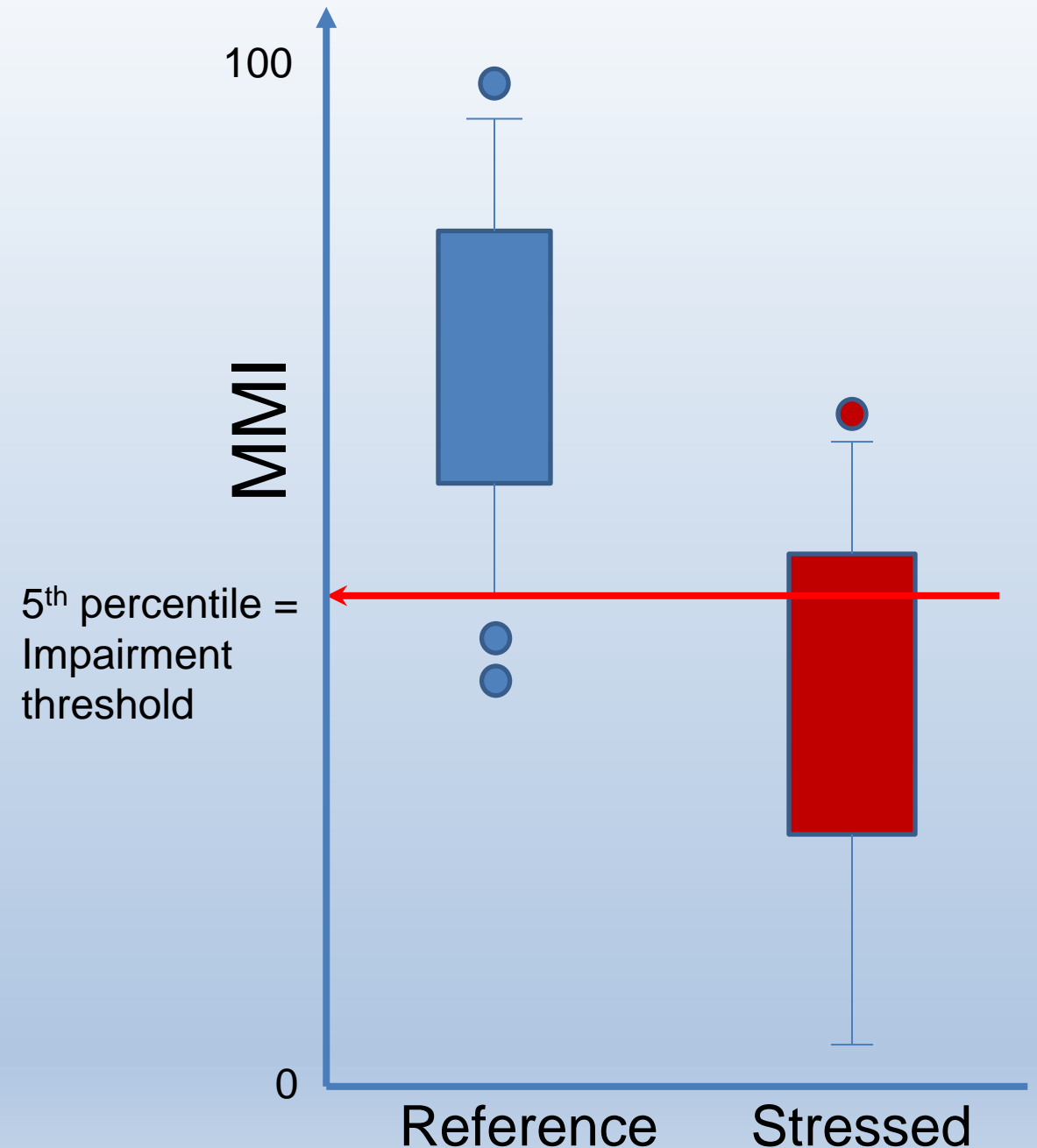
Translating MMIs?

- Multiple levels of aggregation separating raw data and MMI

- Taxa-specific information critical to diagnosing impairment

- "Mapping" taxa change points may serve as translator back to biodiversity losses

- Potential validation of impairment thresholds

Family-level taxonomic data

Genus/OTU level taxonomic data (fixed counts)

?

Aggregate 1

Aggregate 2

Aggregate 3

Aggregate 4

Rescale 1

Rescale 2

MMI
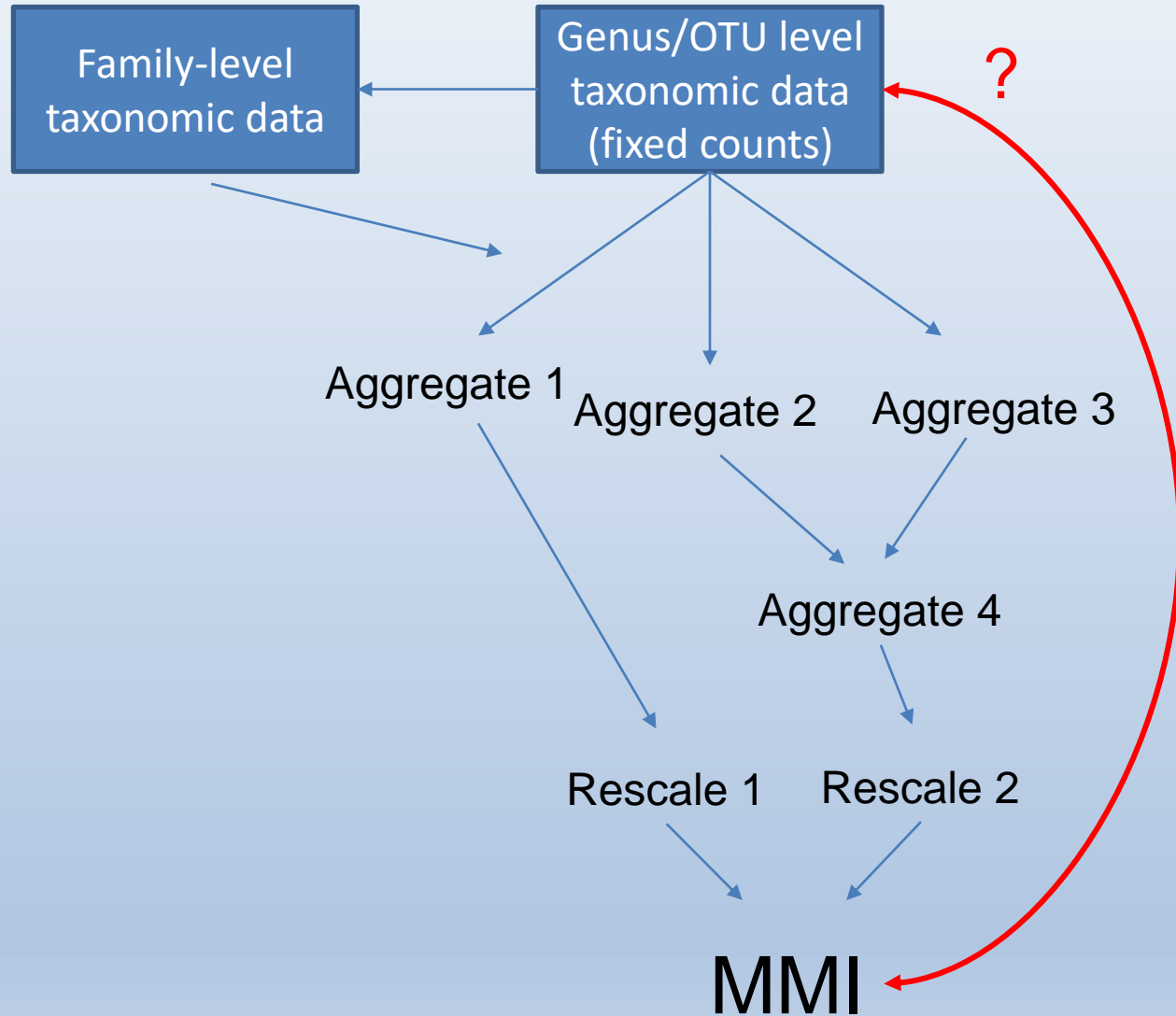
Questions

- Do incremental decreases in MMI correspond to incremental biodiversity losses?

- How well do taxa change points correspond to MMI impairment thresholds?
  - Should they?
  - If not, what does "biological integrity" mean in the context of biodiversity?

MMIs

- **West Virginia Stream Condition Index (WVSCI)**
- Family-level MMI
- Developed using 1996-97 data (67 ref sites)
- Impairment (non-attainment) determined by this index only
  - >72 = attainment
  - 61 = "impaired (maybe)"
  - <50 = "definitely impaired"

- **Genus-Level Index of Most Probable Stream Status (GLIMPSS; Pond et al. 2011)**
- Developed using 3731 sites (391 ref sites)
- Two ecoregions (MT, PL)
- Two seasons (Spr, Sum)
- Different impairment thresholds by ecoregion-season combination (n=4)

Data

- WVDEP (1998-2018)
- Filtered database:
  - Genus-level benthic macroinvertebrates, 200 fixed count
  - RBP, single habitat, streams
  - Survey type: WAP, Random
  - Summer index period
  - Mountain region
  - Catchments < 60 sq. mi.
  - No duplicates, no repeat sites
- Final data set = 1282 sites
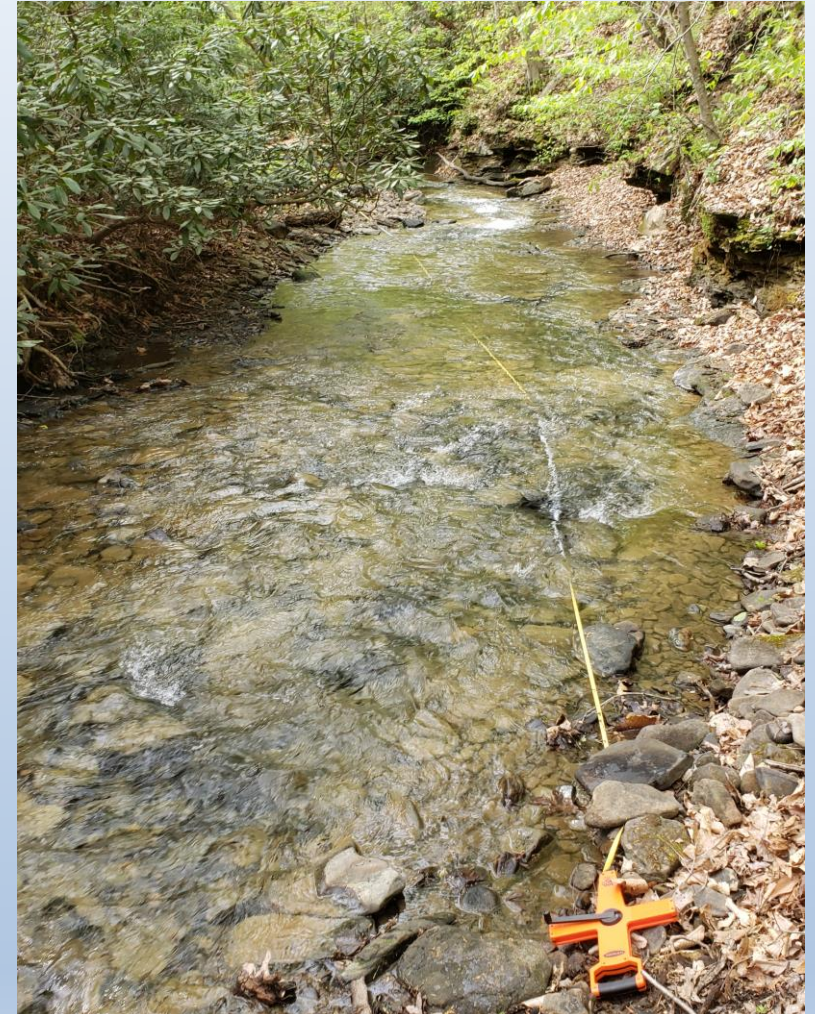
Mapping taxa change points on MMIs

- Threshold Indicator Taxa Analysis (TITAN)*

- Identifies point of largest change in frequency & abundance

- Distinguishes positive (red) and negative (blue) responses

- Evidence for assemblage-level threshold assessed by synchrony in individual taxa change points (Sum(z))

*Baker, ME; King, RS. 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. *Methods in Ecology & Evolution* 1:25-37
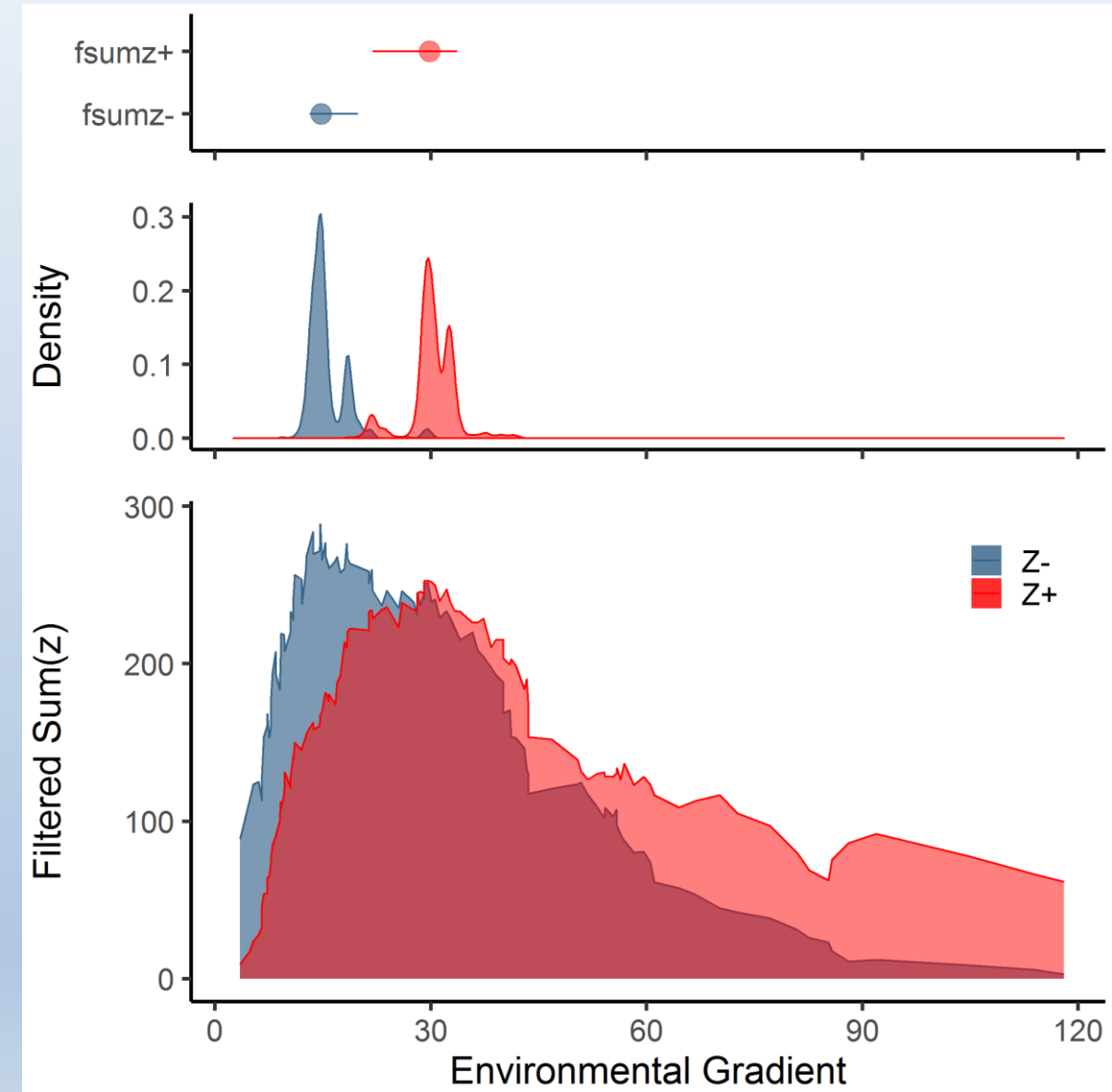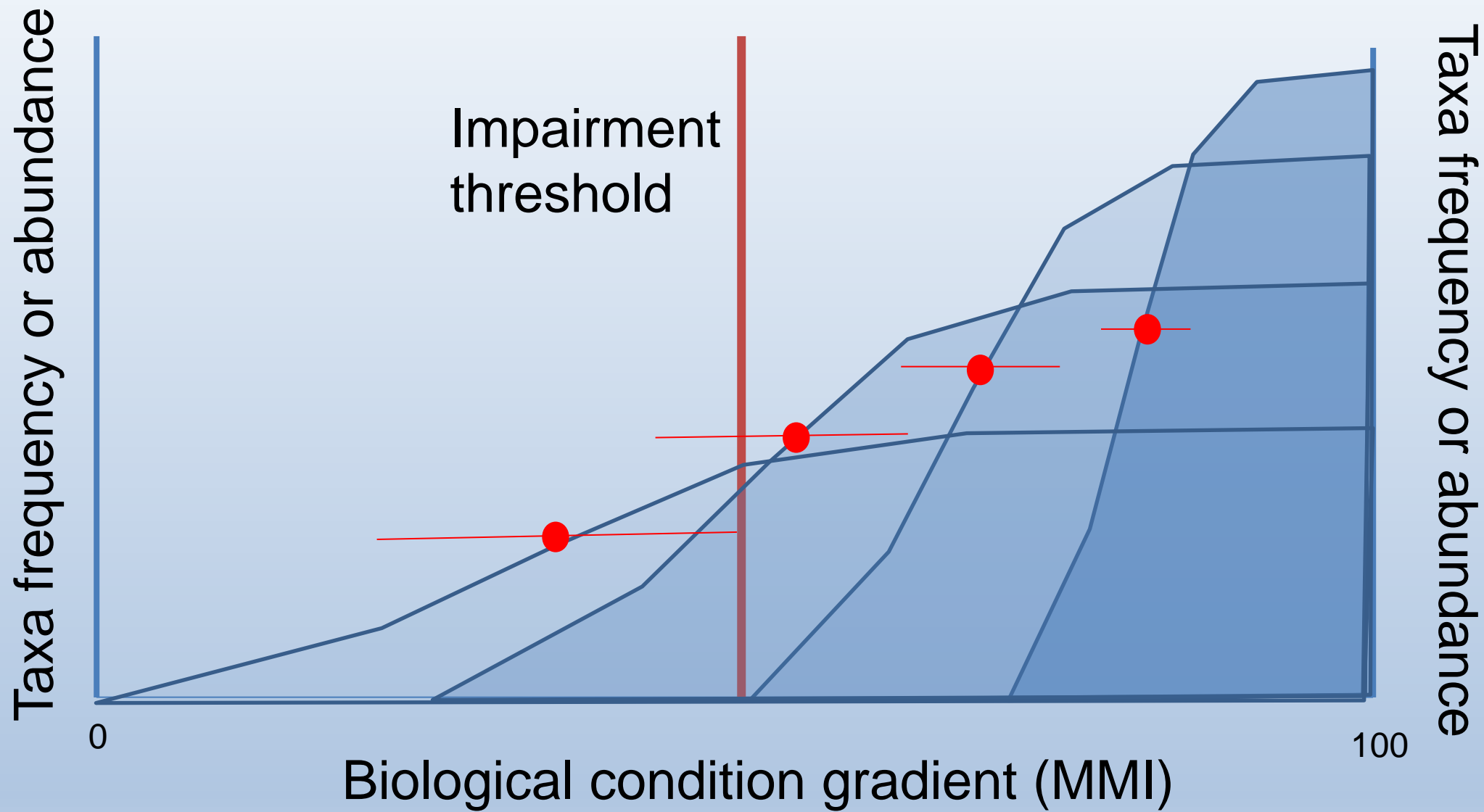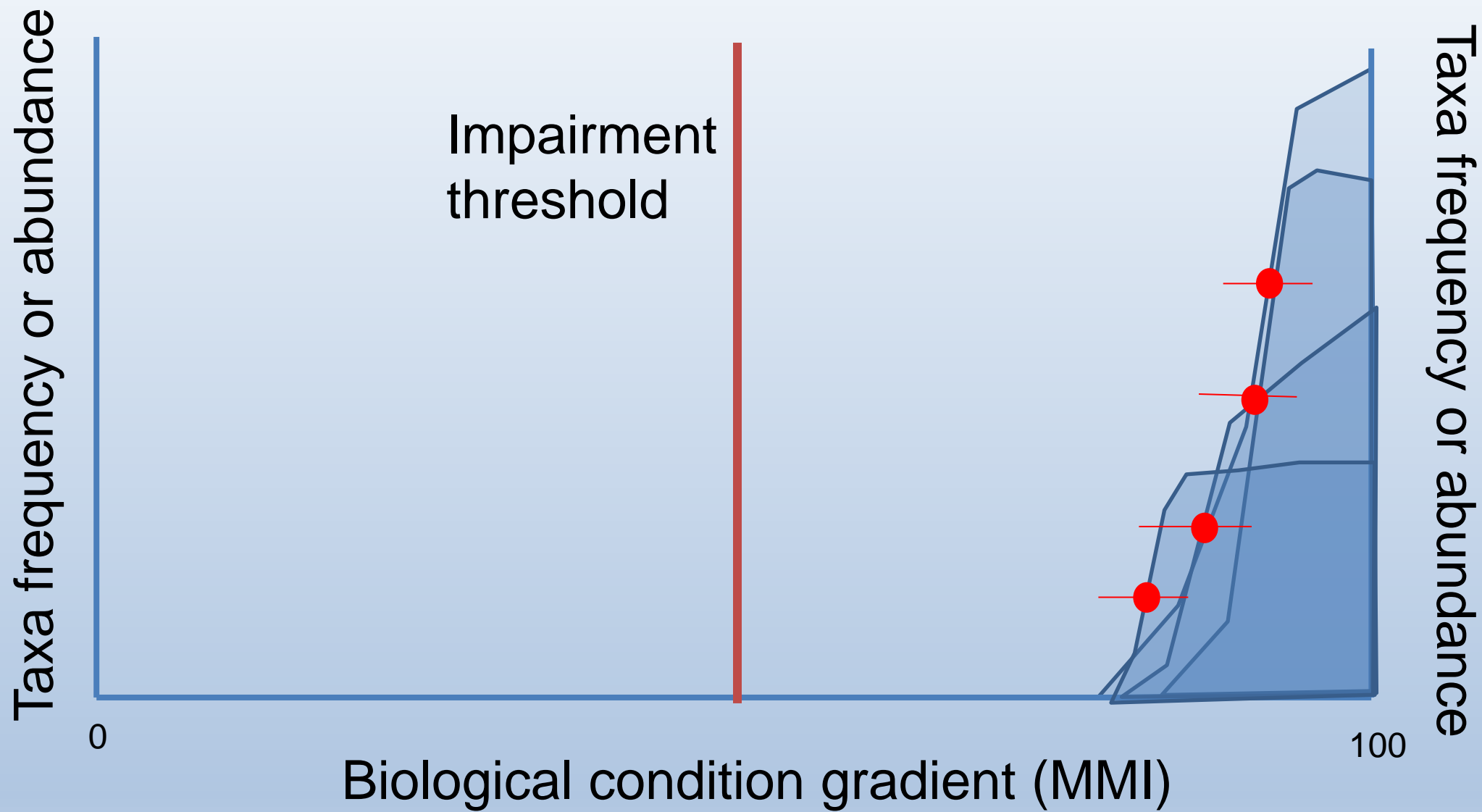
Genus data on WVSCI

136 of 185 genera/OTUs were pure & reliable indicators (74%)

77% of sensitive taxa (85) declined @ WVSCI 80-95
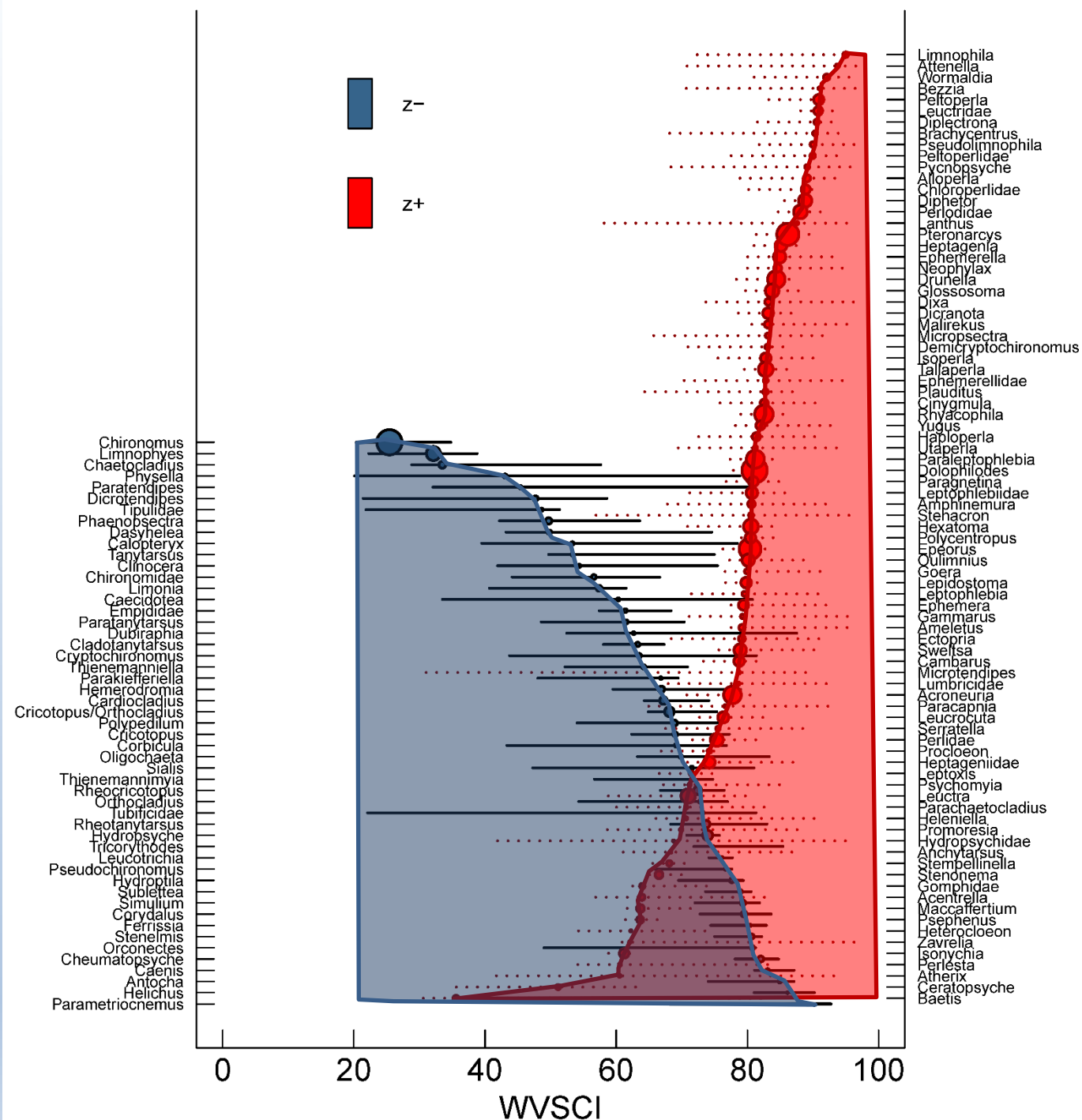
Genus data on WVSCI

Attainment threshold (72) falls below 80% of sensitive taxa change points

"Impairment (maybe)" threshold (61) falls below virtually all sensitive taxa

"Definitely impaired" threshold (50) is definitely impaired
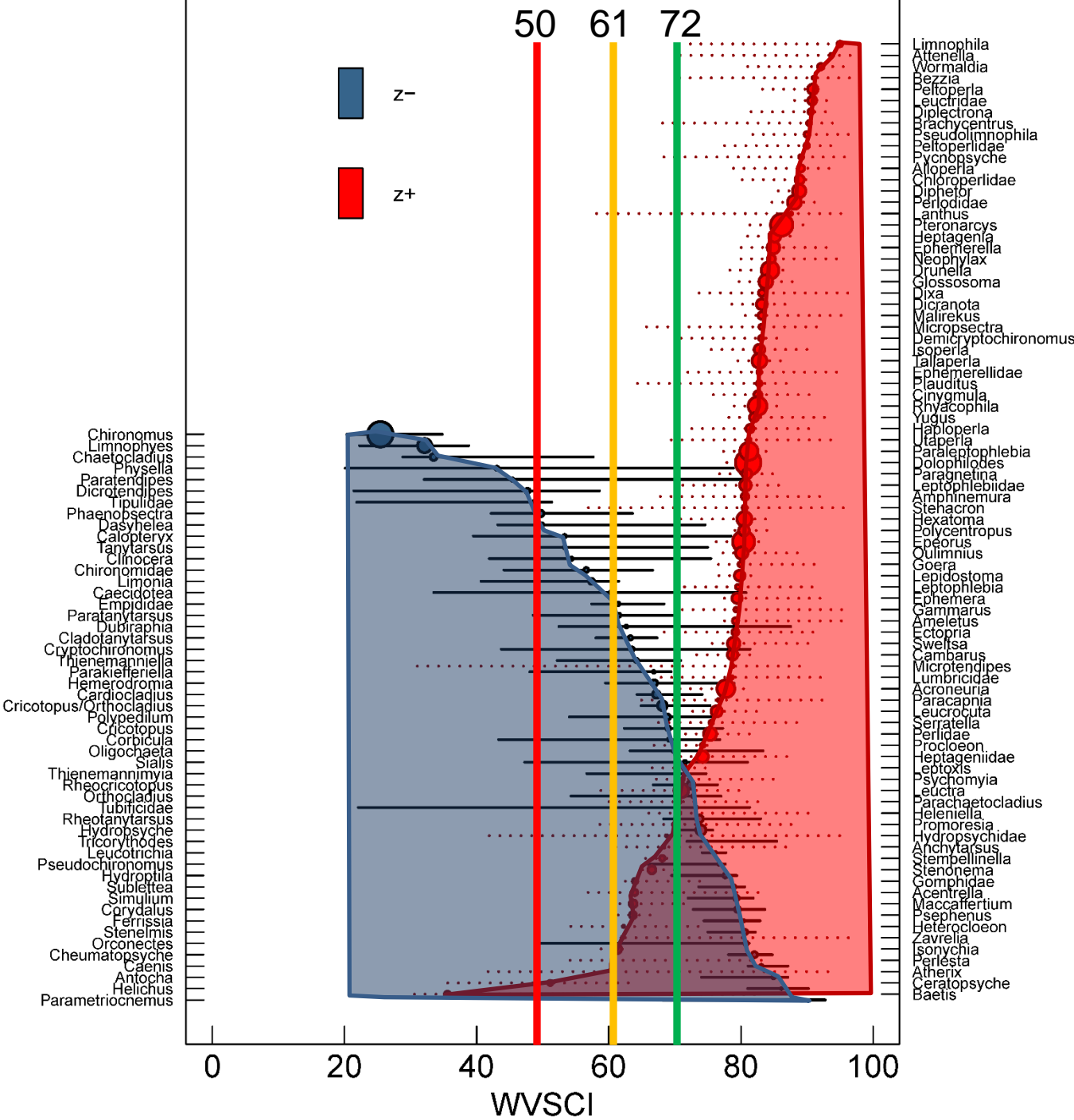
Genus data on WVSCI
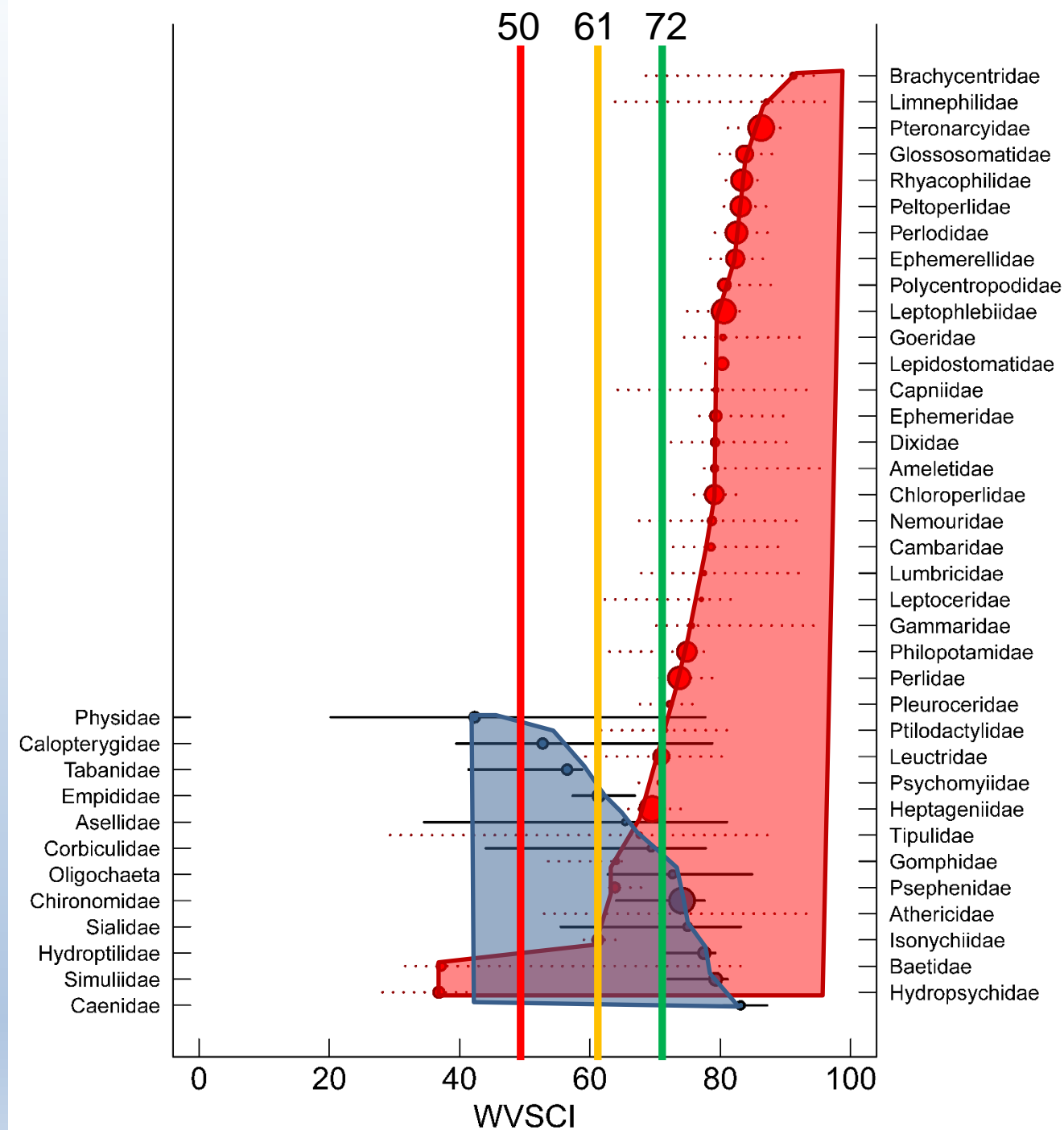
Sensitive taxa change points peaked at WVSCI=79 (77-82)

Tolerant taxa peaked at WVSCI=72 (62-78)
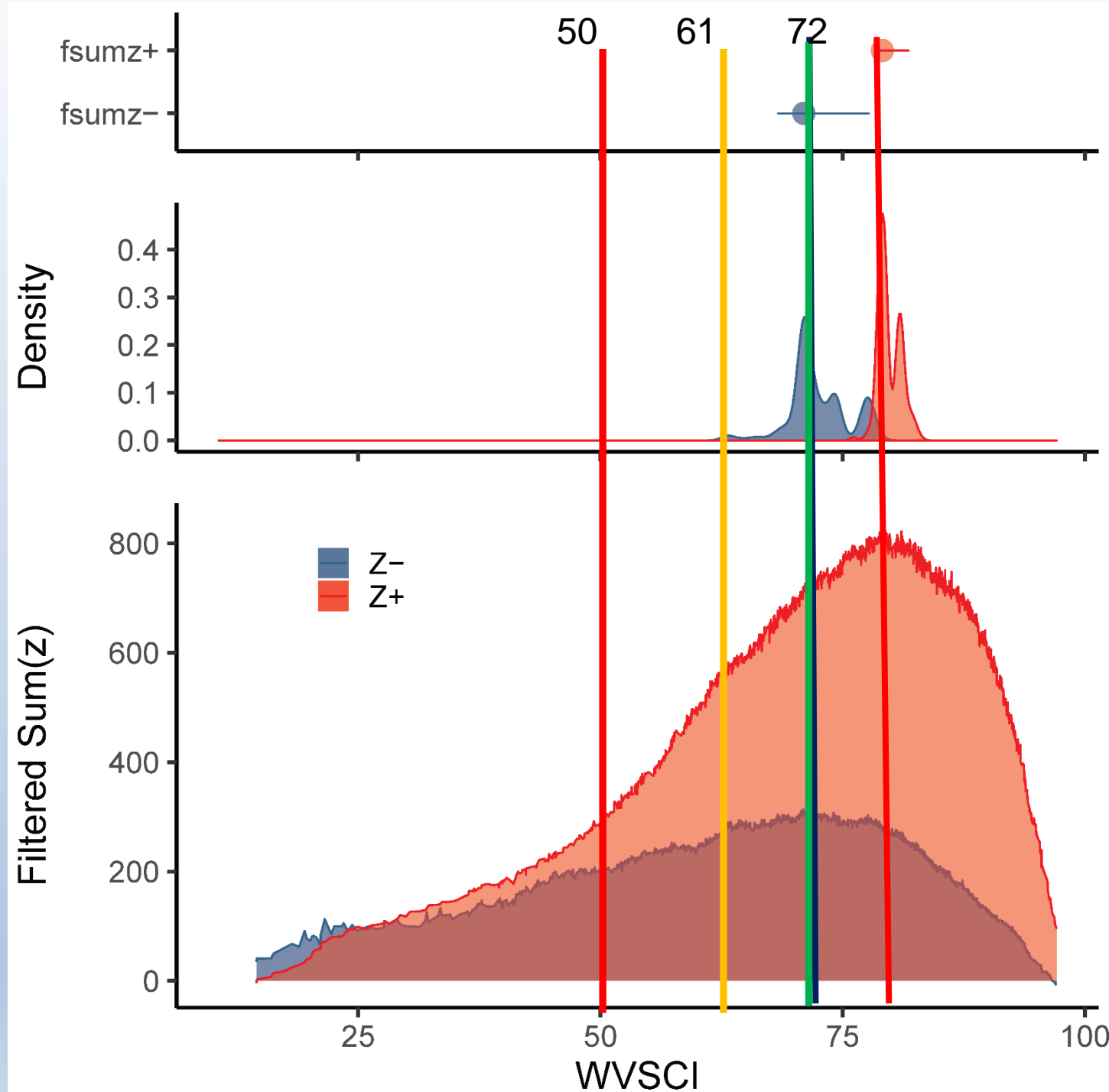
Family data on WVSCI

48 of 70 families/OTUs were pure & reliable indicators in response to WVSCI (69%)

65% of sensitive taxa (36) declined @ WVSCI = 80-95

Family data on WVSCI

Family similar to genus-level results.

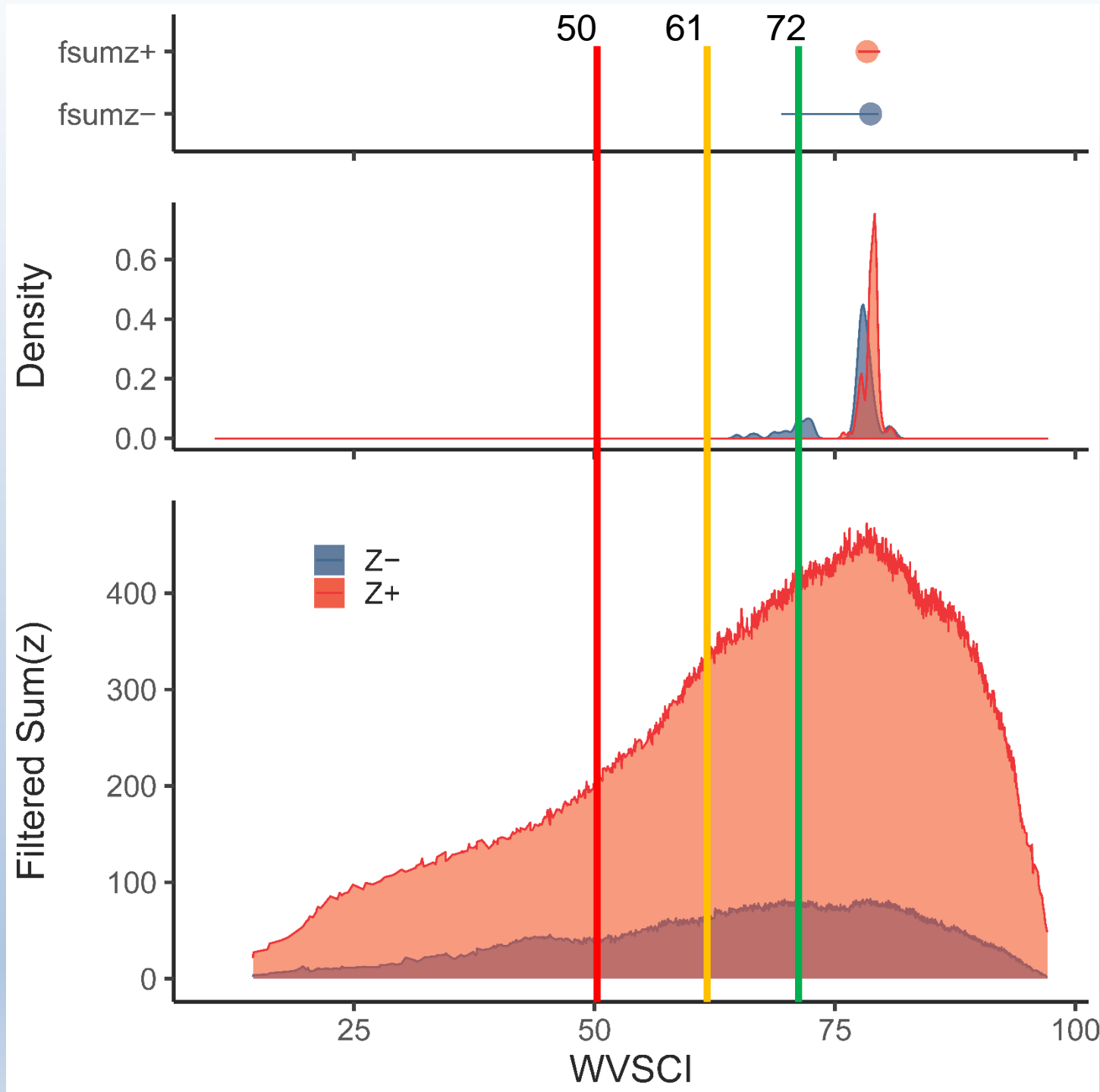Sensitive taxa peaked at WVSCI=78 (77-79)

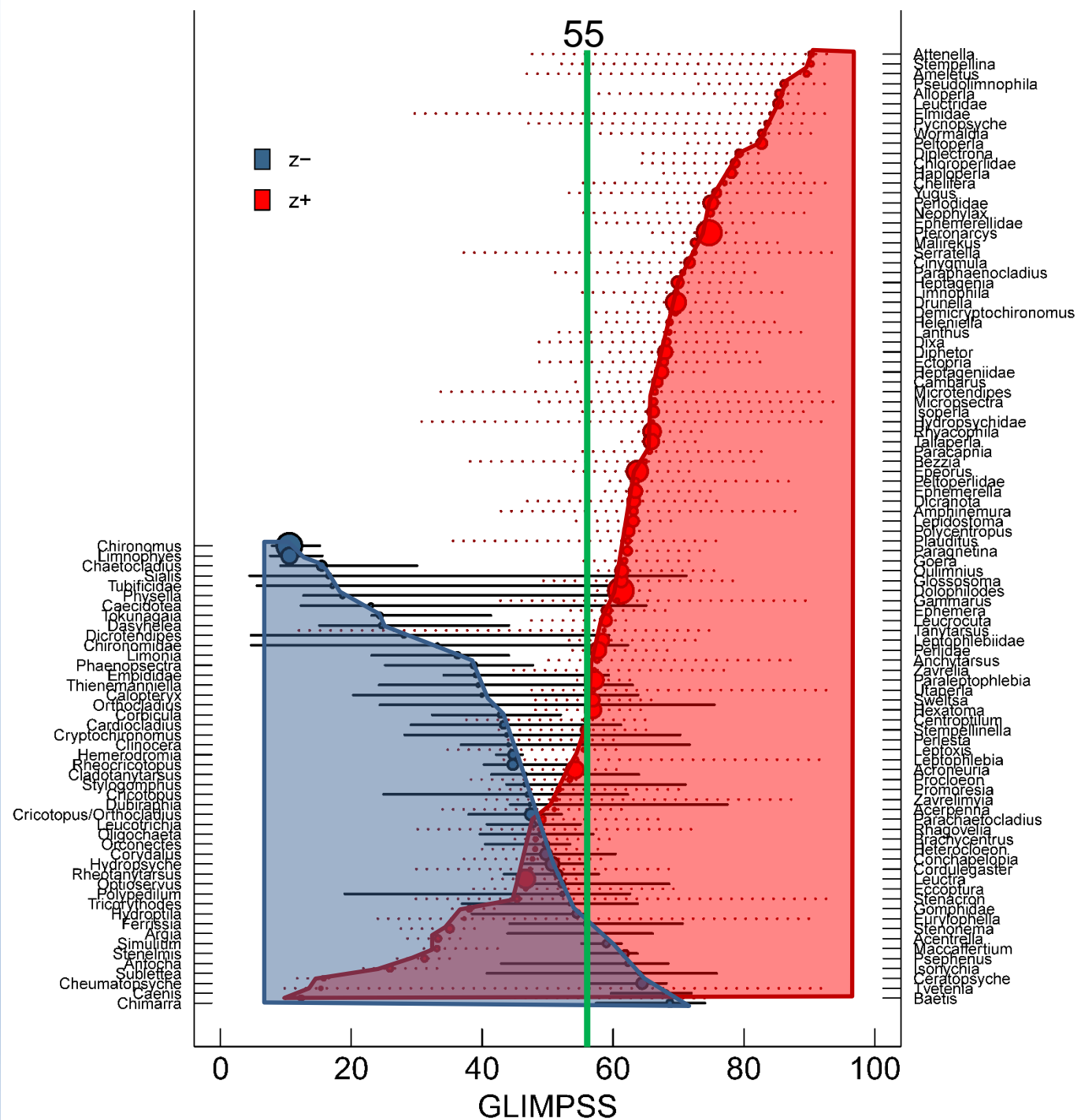Tolerant peak much closer to sensitive at family level, WVSCI=78 (69-79)

Genus data on GLIMPSS

143 of 185 genera/OTUs were indicators in response to GLIMPSS (77%)

Greater continuum of change points on GLIMPSS vs. WVSCI
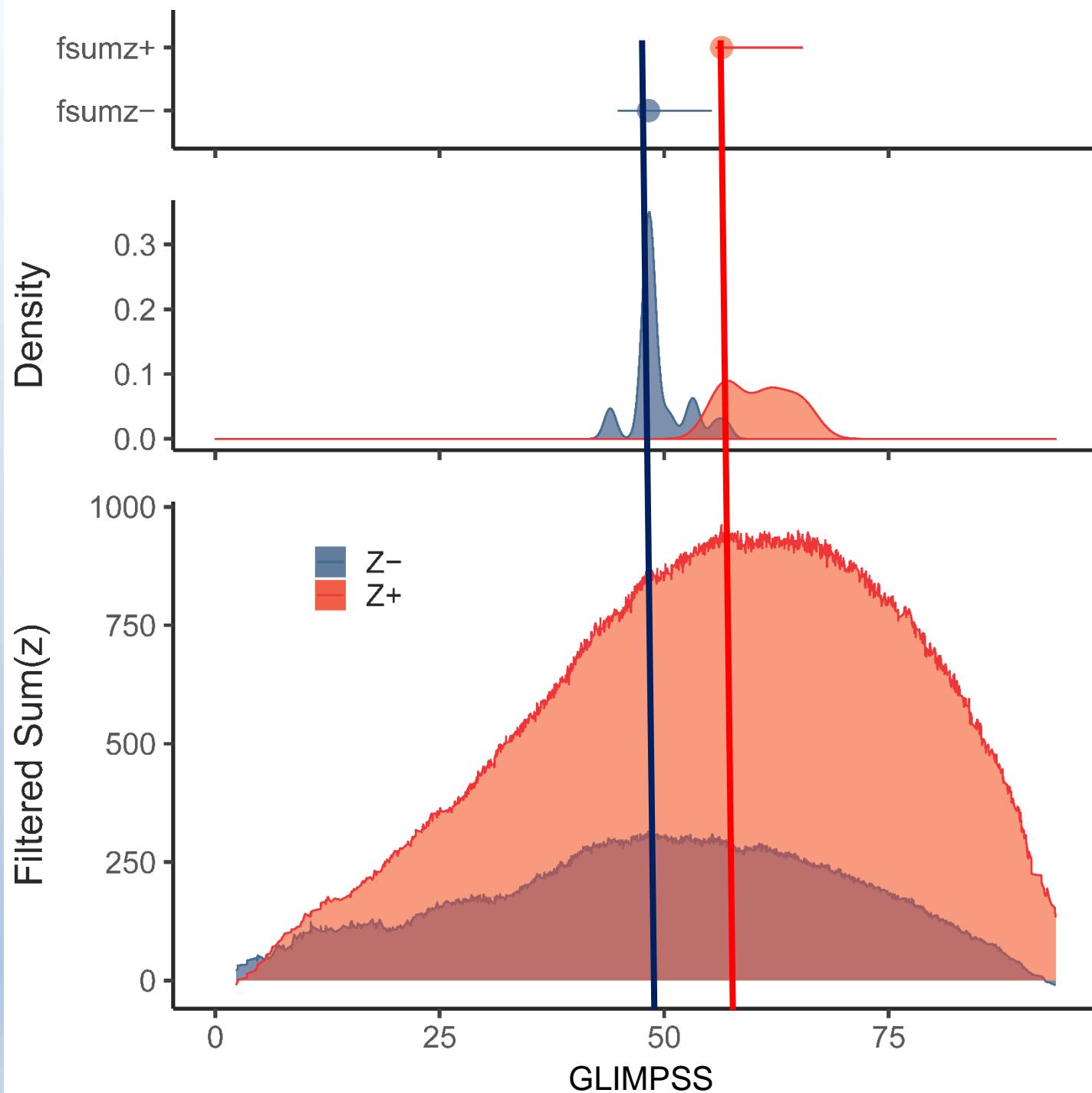
~30 sensitive taxa fell below impairment

Genus data on GLIMPSS

Larger range of sum-z maxima for sensitive taxa compared to WVSCI

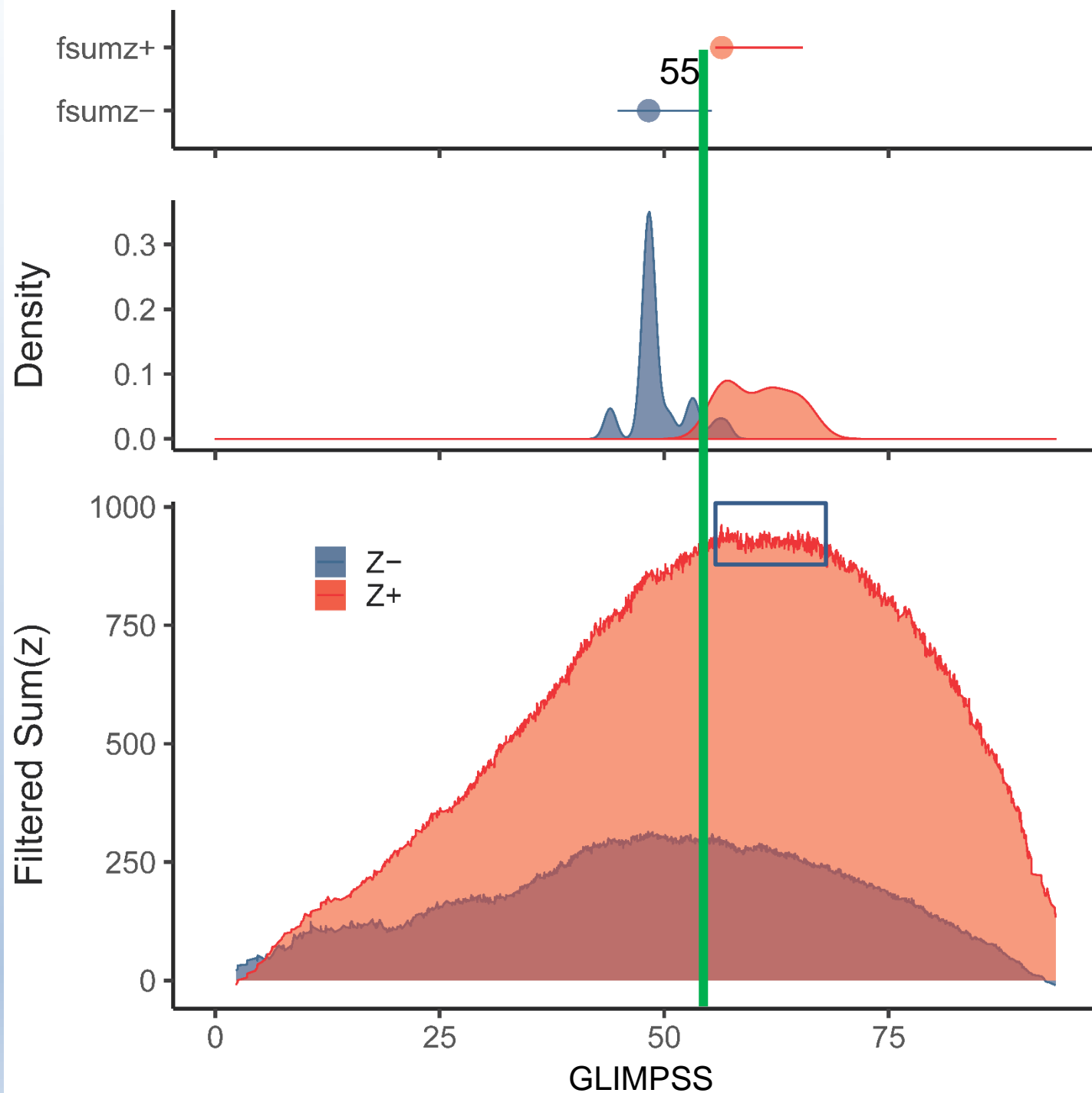Sensitive peak: GLIMPSS=56 (55-65)

Tolerant peak: GLIMPSS=48 (45-55)

Genus data on GLIMPSS

Impairment threshold within 90% CIs of sensitive and tolerant taxa thresholds

GLIMPSS impairment generally supported by TITAN….but note Sum(z) plateau (insert)

Conclusions

- MMIs retain a tremendous amount of signal in individual taxa frequency and abundance

- Mapping taxa change points onto MMIs has potential to effectively translate biodiversity losses along the biological condition gradient

- "Unpacking" MMIs back to taxa abundances may be a useful approach for refining metrics and validating impairment thresholds

Acknowledgements

- David Kahle

- Appalachian Mountain Advocates

- WVDEP